PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : G06F 15/36

A1

(11) International Publication Number: WO 93/01554

(43) International Publication Date: 21 January 1993 (21.01.93)

(21) International Application Number: PCT/AU92/00286

(22) International Filing Date: 16 June 1992 (16.06.92)

(30) Priority data:

PK 7023	2 July 1991 (02.07.91)	AU
PK 9175	29 October 1991 (29.10.91)	AU

(71) Applicant *(for all designated States except US)*: THE ROY MORGAN RESEARCH CENTRE PTY. LTD. [AU/AU]; 411 Collins Street, Melbourne, VIC 3001 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : RENNIE, George, C. [AU/AU]; 50 Nicholson St., Fitzroy, VIC 3065 (AU). GRANT, Peter [GB/AU]; 626 Park Street, Carlton North, VIC 3054 (AU).

(74) Agents: BRETT, Noel, T. et al.; Griffith Hack & Co., 509 St. Kilda Road, Melbourne, VIC 3004 (AU).

(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CS, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, PL, RO, RU, SD, SE, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.

(54) Title: REACH AND FREQUENCY ESTIMATION FOR MEDIA

(57) Abstract

The reach and readership contact frequency are determined for a set of advertising media vehicles (m) with a specific number of advertisements (n_m) per medium. Firstly a subset of survey respondents for each vehicle are selected. The resulting response database is then filtered. Beta distributions are used to build an array containing the probabilities of being exposed to i out of n_m advertisements; the beta distribution parameters being a function of the regularity of exposure of a survey respondent to a media vehicle. The array of probabilities is then combined with survey data on media usage by the respondents to determine a joint frequency distribution of opportunities to see an advertisement, distribution parameters, reach and average contact frequency.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

REACH AND FREQUENCY ESTIMATION FOR MEDIA

TECHNICAL FIELD

The present invention relates to media research, and particularly to an advertising schedule evaluation technique known as 'reach and frequency' estimation.

BACKGROUND OF THE INVENTION

The extent of exposure to mass media is a matter of interest and concern to the owners or publishers of media, to those who use the media to convey a message, normally in the form of paid advertising, and to the agencies they employ to assist them.

These groups are interested in both the size and the nature of the audience. However, absolute information on these questions is not available from any source. In one instance, newspaper and magazine circulation may be audited by an independent body, but this gives only the number of copies sold or distributed and does not indicate how many people or what kind of people read these publications.

To estimate the sizes of the audiences of individual media, it is therefore customary to conduct sample surveys in which a cross-section of the population (or sometimes of a specific subset of the population) is investigated in detail.

There are a number of ways in which the information may be collected and the choice will depend on the objectives of the particular study and on practical and financial considerations. Among other methods, members of a sample may be questioned about their exposure to media within a specific time period, record their media exposure over a period of time, or have their exposure

automatically monitored.

A particular reason for the collection of this information is the assessment (prospective or retrospective) of the performance of alternative media as vehicles for advertising. Advertisers wish to know what they are likely to get, or did get, for their expenditure in terms of the quantity and quality of audience. Media owners, apart from assessing their strengths and weaknesses, wish to impress upon advertisers the benefits of their products.

The average level of exposure to a medium can be readily obtained from these studies. This level is typically taken as the proportion of those surveyed claiming or being observed to see, hear or read the medium at any given point in time or within a given period. For some media it may be necessary to collect information from survey respondents at specific points in time; for others, it may be preferable to spread the data collection over a long period so that short-term fluctuations are evened out in the final aggregated results.

In many cases, media owners and advertisers wish to estimate the potential of two or more media, or of two or more advertisements in one medium, or (most frequently) of combinations of different numbers of advertisements in more than one medium. This requires knowledge both of the cross-usage of distinct media and of the turnover of users of individual media over time. Few individuals are exposed to only one medium and few, if any, are consistent in their usage of any medium.

For instance, people tend to be fairly regular in their reading of a daily newspaper, but even regular readers miss the occasional issue, and there are those who may only read an issue from time to time, or may regularly read it on one particular day only. General interest magazines tend to have a higher proportion of occasional or casual readers. Similarly, many people read one daily

paper, but some habitually read more than one, some sometimes read more than one and some generally read none at all.

The regularity of the exposure of individuals to a medium can be quantified in a number of ways. 'Turnover' and its complement 'loyalty' are two. In Australia, the most common measure is 'casualness'. The casualness (γ) of a medium is defined as the ratio of the additional reach of a second issue (over the first) to the additional reach which could be expected if the persons reading, seeing or hearing the first and second issues were chosen independently of one another. Thus, 0% casualness implies that people always do the same thing, whilst 100% casualness constitutes completely random behavior.

Media surveys must in one way or another assess the degree of consistency of usage of individual media and the extent of cross-usage. Where the behavior of respondents cannot be studied over a prolonged period, these must be estimated. Two methods of doing this are commonly used. Respondents are asked to generalise their behavior in respect of individual media to give some indication of the regularity of their exposure to them, or they are re-contacted at a later time and asked about the same media, thus providing an indication of audience turnover. Such methods are used to provide indications of the consistency of use of individual media within the overall population and within designated subsets of it.

The information provided by these studies is used to simulate or assess the ability of alternative advertising schedules (a schedule being a list of media vehicles and numbers and types of advertisements inserted in them) to reach designated segments of the population. The principal measures commonly used are: the number or proportion of people exposed to at least one of the advertising insertions, the total number of impacts (where an impact is defined as one person being exposed to one

insertion) and the average impacts amongst those reached (which equals total impacts divided by reach). These are generally referred to as 'reach', 'total impacts' and 'average frequency' respectively. In addition, it may be desirable to estimate the numbers exposed once, twice, three times, etc.; this is commonly known as the OTS (opportunities-to-see) distribution or the 'number seen' distribution. These two expressions are equivalent and may be used interchangeably.

DESCRIPTION OF THE PRIOR ART

There are two types of media reach and frequency system in general use. These utilize either

(i) global formula methods or

(ii) personal probability methods.

Global formula methods, as the name implies, are based on the application of a formula or an algorithm to aggregated survey data such as average readerships and pairwise cross-readerships, in the case of print media.

Personal probability models, on the other hand, treat each individual in the database separately, with each respondent assigned a probability of seeing each medium. The probabilities of 0, 1, 2,..., n OTS are then evaluated using the binomial distribution and these are then aggregated over all respondents to give the OTS distribution for the total population.

Both techniques have their drawbacks. Critics of global formula methods cite:

(a) The problem of declining reach, whereby an extra insertion can cause the estimate of the overall reach of a schedule to decline;

(b) Non-additivity, whereby the distributions for complementary subsets of a population (*e.g.* men and women) do not necessarily reconcile with the distribution for the total (*e.g.* all people);

(c) The inability to model correctly non-overlapping combinations of media (*e.g.* in schedules across geographical regions that are not covered by all media);

(d) The failure to model the actual shape of the OTS distribution (*e.g.* in cases where a high-audience medium is included in a schedule with a low-audience medium); and

(e) Order dependency, whereby the answer given may be dependent on the order in which the media are specified.

To a very large extent the success or failure of personal probability methods depends on the adequacy of the determination of the personal probabilities. Critics of these methods draw attention to the following drawbacks:

(a) Failure to reconcile with 'head-count' data among all subsets;

(b) The dilution of cross-media relationships; and

(c) The underestimation of reach.

Presently available systems are also liable to encounter problems in dealing with incomplete data. In some cases, where a survey covers an extended period, information may have been collected on particular media vehicles for only part of the time, either because of a change in the scope of the survey or because those media were only available for part of the period. In some cases, it is possible to process a schedule including one such incomplete medium by applying a weight to the aggregate results relating to that medium, but where two or more such incomplete media are included, this is not possible.

Where a survey covers an extended period, it may be desirable to restrict the evaluation of a schedule to respondents interviewed during only a specific portion of the overall time period. In such cases, the resulting

estimates of numbers exposed are depressed unless a scaling method or an alternative set of respondent weights is employed which grosses the reduced sample up to the required population.

SUMMARY OF THE INVENTION

The present invention is directed to overcoming some or all of the drawbacks listed above.

Therefore, as a non-limiting statement to indicate the scope and intention of the invention, it can be said, in one instance, that the invention provides an estimation method for an information processing system including a processor means, a memory means coupled to the processor means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means. The estimation method comprises the steps of:

(A) selecting, by the input means, at least one schedule of media vehicles (m) and, for each selected schedule, specifying a number of insertions (n_m) in each media vehicle;

(B) defining, by the input means, at least one filter specifying a set of the survey respondents for which each schedule is to be evaluated;

(C) applying, by the processor means, the filter to the database;

(D) for each media vehicle, calculating, by the processor means, an array of the probabilities of being

exposed to i out of n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle;

(E) combining, by the processor means, the data relating to the usage of each selected media vehicle with said array of probabilities to give a probability of being exposed to i insertions in each selected media vehicle, combining the probabilities of being exposed to i insertions across the media vehicles and accumulating over all respondents passing the filter to yield a composite distribution (X_i);

(F) by the processor means, within the set of survey respondents defined by the filter:

(i) summing media vehicle usage to give total estimated usage for each selected media vehicle, and summing the media vehicle usage of each pair of media vehicles (m_1, m_2) to give total estimated cross-usage of each pair of media vehicles,

(ii) estimating the mean of the true OTS (opportunities-to-see) distribution for each selected schedule from said total estimated usage for each selected media vehicle,

(iii) estimating the variance of the true OTS distribution for each schedule from the said mean, the said total estimated cross-usage and the said estimates of regularity of exposure for each media vehicle;

(G) operating, by the processor means, on each said composite distribution (X_i) to modify it so that it matches

with the said mean and the said variance of the true OTS distribution to form a final composite OTS distribution; and

(H) outputting, by the output means, the final composite OTS distribution for each selected schedule.

The method can preferably calculate the reach for each selected schedule by forming the sum of the final composite OTS distribution frequencies for the cases of insertions being one and more. Alternately, the reach can be calculated by taking the complement of the final composite OTS distribution frequency for the case of the number of insertions being zero. The average frequency for each selected schedule can then be determined as the total impacts divided by the reach.

The method can also determine a partial final composite OTS distribution of one or more terms. In this case, the reach is determined as the complement of the distribution frequency (X_0) of zero exposures.

The invention also provides for an estimation method for an information processing system as described above, which method comprises the steps of:

(AA) determining, by the processor means, whether any adjustment is to be made to allow for

(a) media vehicles not included in the survey throughout the whole survey period;
(b) sample subsets defined wholly or partly in terms of time periods;
(c) sample subsets defined in terms of information not collected throughout the whole survey period;

(d) any sample subset used at any stage in the calculation which is formed as the intersection of any two or more sample subsets of the types (a)-(c);

(BB) calculating, by the processor means, the adjustment factors for

(a) the total estimated usage of each media vehicle;
(b) the total estimated cross-usage of each pair of media vehicles;
(c) the frequencies (X_0, X_1,..., X_N) of the composite frequency distribution; and

(CC) applying, by the processor means, these adjustment factors to produce revised estimates of the composite distribution frequencies, total impacts and the mean and variance of the true OTS distribution.

The invention also contemplates an information processing system comprising a memory means, input means, a processor means and output means operable in accordance with the method as described in the foregoing paragraphs. Such a system may preferably also include a printer for printing the final composite OTS distribution(s).

BRIEF DESCRIPTION OF THE DRAWINGS

In order that the invention can be more clearly explained, an embodiment will be described with reference to the accompanying drawings in which:

Figure 1 shows a flow diagram of the method of the invention;

Figure 2 shows a hardware realization of a system embodying the invention suitable for a personal computer

environment; and

Figure 3 shows a hardware environment similar to Figure 2, but suited to a mini-computer or mainframe.

DESCRIPTION OF PREFERRED EMBODIMENTS

An illustrative embodiment of the present invention relates to print media; however, it is to be appreciated that the invention is equally applicable to other media or to cross-media implementations, *e.g.* print in combination with television or radio. In this regard, the expression 'media vehicle' is to be taken as relating to any appropriate forum in which advertising could be placed.

Therefore, when 'media vehicle' is used in the case of print media, it refers to any individual 'publication'. In the case of television and radio, it typically designates individual programs or defined time periods, etc.

The term 'insertions' is used to encompass a unit or instance of an advertisement in any one media vehicle. Therefore, in print media an insertion may represent one advertisement placed in one issue of a particular publication, whereas for television, an insertion may represent one advertisement during a particular program or time slot. Similar considerations apply for radio and cinema.

The embodiment will be described as a system that is conveniently realized by use of a suite of computer files designed to run on a personal computer (operating independently or attached to a network) or on a larger mini-computer or mainframe computer. The suite of programmes is commercialy available from the applicant company and sold under the name ASTEROID.

Figure 2 shows, in block diagram form, personal computer-based elements which can be utilized to implement the system. Memory means is provided variously by the

internal memory block (such as RAM), the hard disk (external memory) block, the floppy disk block and the virtual disk block. The virtual disk is extended internal memory which may assist in speeding up calculations. The floppy disk is optional, but it does provide a useful facility for inputting the survey database or the host software. Processor means is shown by the processor block. Input means is represented by the keyboard block. This allows a user to input the user specified criteria as will be presently discussed. The output means can be such as the VDU block for a visual representation, the printer block to which hard copy can be directed or a mass storage medium such as the hard disk on which output can be stored. The elements identified by an asterisk are considered optional to a minimal system, but would most often be provided.

It has been found that the following specification operates satisfactorily: an IBM or compatible PC (type XT or upwards), having at least 640 kb of memory (RAM). The environment/operating system should be MS-DOS/PC-DOS (or equivalent) version 3.0 or later, with the screen driver ANSI.SYS installed. A numeric (math) co-processor is also advantageous in speeding up calculation times, as is extended memory.

Figure 3 is similar to Figure 2, but is a block diagram representation of a mini-computer or mainframe system. The internal memory block, disk (external memory) block and tape block are each examples of memory means. Input means is provided by the ANSI standard terminal. The terminal has a screen, which together with the printer block form the output means. Again, processor means is realized by the processor block. Typically, a VAX or Microvax system running VMS 5.0 or later is acceptable.

The hardware described for supporting the system is exemplary only. The system has been designed to be easily transportable; hence many other hardware platforms could

equally be utilized.

The system incorporates database files pertaining to survey data from some number of respondents, as well as executable programs, and when run, the system outputs a tangible written or printed presentation of the OTS distribution together with the reach and average frequency estimation.

There are five basic types of file stored on hard disk:

1. The main program file.
2. Local files (files specific to the user and not available to other users). These include files describing the configuration of the user's preferred output format, private dictionary files, input and output files generated by the user, etc.
3. Data files. These include the main database file, the files governing access to it and the file of weighting factors. In the case of media surveys there is an additional 'respondent loyalty' file.
4. User utilities. These assist in customising reference files and in the creation of private dictionaries.
5. Reference files. These are accessible to all users and include the standard (or 'public') dictionary files, files containing the menus, error and information messages and prompts.

The main database file contains, in compressed form, the respondent survey information (demographic, media exposure, product usage, attitudes, etc.). Its generation is described below.

The data from which the system compiles its results may be taken from conventional market research questionnaires, from active or passive metering or recording devices, or by other means. Because of the nature of many of the questions, which allow (indeed encourage) multiple responses (*e.g.* "Which of these ...

have you ever used?") it may be expedient to record the information as though by physically punching holes in a punch card, although this is not an essential strategy.

The data from a punch card is essentially binary, that is, one bit can represent each punch position. Normally a physical record in a file represents the contents of one nominal card. There are various conventions which can be used.

The punch card 'used' in the present system is an IBM standard card on which the data area is divided into 80 'columns' and 12 'rows'. Thus there are 960 possible punch positions on a card. Data from one respondent may require more than one card and therefore, any piece of information is identified in terms of its card number, column number and row (or code) number. The system uses a notation with, for example, card 2, column 33, code 6 being written as:

#2 /33 6

Each column covers twelve rows or codes. These are designated V, X, 0, 1...9. 'V' is the position nearest the top edge of the card and 9 is the bottom position. The system assumes the logical sequence stated above.

The software takes binary files in which the record represents some or all of the information about one survey respondent and 'inverts' this into files in which each logical record represents one punch position. The result is a string of bits which indicate whether each respondent has or has not a particular characteristic (maleness, reading the Washington Post, owning a dishwasher, etc.). Because only certain facts, forming generally a very small proportion of the total data available, are required to produce each table, extracting only the records containing the relevant information minimises or eliminates redundant input/output.

The inverted file is compressed by the omission of records relating to unused punch positions. Because the logical records may be too long to handle comfortably, they are broken up into segments. The file is scrambled to make it difficult to read by any other program. This file is always assigned the '.CPR' file-type. A pointer file contains values which allow the system to calculate the physical location of any segment required.

The respondent loyalty file is a direct access file which contains loyalty estimates p_m for each respondent for each medium m covered by the survey. For convenience in the construction of lookup tables, these take on a finite set of discrete values (typically 256) and are encoded as characters.

Weighting factors may be assigned to each respondent. These factors are calculated to redress any imbalances in the composition of the sample and in addition are normally scaled so that the sum of the weights represents the estimated size of a parent population (*e.g.* all households, all persons). In the present embodiment, these weights may be disabled (replaced with dummy weights of 1.0) by the user if desired.

The .CPR file is accessed by specifying the card locations relating to the data to be extracted. The user, however, normally specifies these as a data name, which is associated with a definition in primitive (card location) terms. The data name specified may in fact be a name associated with a series of related data names. Thus a user may ask for 'men' (one subset of the sample) or 'sex' (which will yield 'men' and 'women').

Names like 'sex' are called 'variables'. A variable is composed of two or more 'groups', each representing a specific subset of the survey sample. These groups need not be mutually exclusive and need not collectively represent the entire population.

Subsets of the sample may also be defined without

being a group in a variable. These are referred to as 'entities'.

A filter is a definition of a subset of the data to which an analysis is to be restricted. This definition may be expressed in terms of card locations or data names or a mixture of the two, data names being converted into their existing card-location definitions during processing. Elements of the definition are linked by logical operators to indicate whether the union or intersection of two elements or the complement of an element is required. The term 'filter' is also applied to the process of applying this definition to exclude all data other than that defined from subsequent processing.

Each variable, group and entity is defined in a dictionary file. Each variable has a data name or tag of up to twelve characters and this tag is associated with a text string used to provide a fuller description in the final table. Each group within the variable has an optional tag, an optional verbal description, and a (mandatory) definition, in terms of punch card locations, of the data comprising that group.

This definition may encompass, at its simplest, a single card location. Alternatively, the definition may consist of a series of card locations, logically linked. For instance, 'old men' may be defined as 'any code 7-9 on column 23 of card 1 but also code 2 on column 65 of card 1'. The actual definition would look like this:

#1 /23 7-9 & /65 2

This definition is read and interpreted by the software into a series of steps of reading and logically combining the contents of specific records to form the rows and columns of a table and logically combining each row and column to form the individual cells of the table.

A subset of the sample normally corresponds to a

subset of the population. However, in a survey conducted over a substantial period, a subset defined purely in terms of the date of interview may need to represent the entire population. To gross up from such a subset requires either an alternative set of weights specific to each definable time period or, as in the present case, a mechanism for making the appropriate adjustment.

The program checks the definition to ascertain whether any elements of it relate to time periods. It recognises this through card locations identified when the database was created. If any such time elements are found, a further series of logical steps is generated to extract separately, and where necessary combine, the contents of records relating to the referenced time periods. This information is used to calculate adjustment factors:

(a) for media vehicles not included in the survey throughout the whole survey period;

(b) for sample subsets defined wholly or partly in terms of time periods;

(c) for sample subsets defined in terms of information not collected throughout the whole survey period;

(d) for any sample subset used at any stage in the calculation which is formed as the intersection of any two or more sample subsets of the types (a)-(c) above.

The data relating to each time element in a definition are extracted as bit-strings. Multiple time bit-strings within a definition are combined by logically 'OR'ing (bits in a resultant string are set 'on' if the corresponding bit in any input string is 'on', otherwise set 'off') to yield a single time bit-string representing the time elements. The weighted total of all respondents is divided by the weighted sum of the 'on' bits in the resultant time bit-string to give an adjustment factor for

that subset and for any intersections with other subsets whose definition do not contain time elements.

Alternatively, the adjustment factors may be calculated as the weighted sum of respondents in a filtered subset divided by the weighted sum of respondents in the subset who are also identified ('on' bits) in the time bit string.

Where two subsets with definitions containing time elements intersect (for instance, the definitions of a row and column of a table, or a media vehicle with a filter) their time bit-strings are also logically 'AND'ed (bits in the resultant string are set 'on' if the corresponding bits in all input strings are 'on', otherwise set 'off'.) An adjustment factor is then calculated as above.

When the complete definitions are evaluated, applied, combined and aggregated, these adjustment factors are applied to the resulting aggregates so that the final output figures are comparable estimates of the required population. Of course, if there is effectively no intersection (for instance, two media vehicles about which no data were ever gathered at the same time) there can be no estimate of joint audience, except by projective or data fusion techniques, but this is a problem of media surveys generally and not confined to the processing system described here.

This feature may be enabled or disabled by the user. When it is enabled, the time corrections are made automatically requiring no intervention by the user.

The main executable program ASTEROID.EXE performs the main analysis function required with survey data, that being two-way tabulations of selected variables or subsets, restricted (filtered) where required to desired subsets. It provides the user with a high degree of flexibility in the selection of data and the presentation of output. For most functions, users can choose courses of action or data either by selecting items from menus or

by entering semi-natural language commands.

In relation to the present invention, the program also can be used to generate an OTS distribution for any number of impacts (*i.e.* either the complete OTS distribution or a partial or truncated OTS distribution can be output containing any number or combination of impacts from zero up to the total number of insertions), the reach and average frequency estimates for each defined filtered population. These require the user selection of the respondent survey data and one or more media schedules and the definition of one or more input filters or subsets of the data over which the schedule is to be evaluated. All this information is input to the system via a keyboard and is viewable by the user on the video screen of a personal or mainframe computer. The input data will also be stored in a user-named local file of the type .SCH.

In specifying the schedules, the media vehicles in which the advertisement insertions are to be considered must first be selected. Each schedule is then completed by specifying the number of insertions in each selected media vehicle. The final step is to specify one or more filters, which typically designate demographic target groups.

In the following example particular to Australia:

NAT-GEOGRAPH
NEW-IDEA
ADV-M-F

are tags for National Geographic, New Idea and the Monday-Friday average issue of the Adelaide Advertiser newspaper. These are respectively a monthly magazine, a weekly magazine and a daily newspaper published in the State of South Australia. Four completed schedules for the three filters of Gentlemen, Ladies and All Adults are shown:

Schedule File: SCH01
Filter: Gentlemen

Publications, m		Schedules (Insertions, n)		
NAT-GEOGRAPH	3	4	2	4
NEW-IDEA	2	3	2	2
ADV-M-F	1	2	2	0

Schedule File: SCH01
Filter: Ladies

Publications, m		Schedules (Insertions, n)		
NAT-GEOGRAPH	3	4	2	4
NEW-IDEA	2	3	2	2
ADV-M-F	1	2	2	0

Schedule File: SCH01
Filter: All Adults

Publications, m		Schedules (Insertions, n)		
NAT-GEOGRAPH	3	4	2	4
NEW-IDEA	2	3	2	2
ADV-M-F	1	2	2	0

Once all the input data have been specified, the program ASTEROID.EXE is able to estimate the OTS distribution, reach and average frequency for each schedule and each filter as a basis for comparison of advertising strategies. The following steps are involved. The adjustment factors to correct for the presence of time-dependent definitions have been omitted from the formulae for clarity.

<u>Step 1</u> The system accepts as input the specification of the schedules including the number of insertions, n_m, for

each medium, m and the filters required.

<u>Step 2</u> For each possible value of respondent loyalty, p_m, the system generates an n_m long array of the probability of the respondent being exposed to 0, 1, 2, ..., (n_m-1) insertions, viz.,

$$P_{m0},\ P_{m1},\ P_{m2},\ \ldots,\ P_{m(n_m-1)}$$

using an individual beta distribution with parameters:

$$\frac{p_m G_m}{1-G_m}\ ,\ \frac{(1-p_m)\,G_m}{1-G_m}$$

where G_m is the internal casualness of the publication. (Whilst the beta binomial distribution is preferred, it is also possible to utilize other discrete frequency distributions of appropriate order in performing this step). The system stores each of the arrays for subsequent use.

In this way, the survey respondent may be viewed not as an individual in the population but as a demographic 'cell', *i.e.* the representation of a number of individuals with the same demographic characteristics. The internal casualness G_m is the within-demographic component of casualness. The between-demographic component is:

$$\sum\nolimits_{pop} p_m(1-p_m)\,/\,[\bar{p}(1-\bar{p})]$$

where the sum Σ_{pop} may be weighted to a parent population if this is required, and the overall casualness is the

product of the two.

Step 3.1 The system then reads the survey database file and simultaneously the respondent loyalty file. For each respondent passing the filter, the system takes the actual readership of each publication δ_m and combines it with the probability distribution array generated in step 2 to give the probability of seeing 0, 1, ..., n_m issues, viz.,

$$\begin{aligned} \text{Probability (seeing i)} &= P_{mi} && \text{if } \delta_m = 0 \\ &= P_{m(i-1)} && \text{if } \delta_m = 1 \end{aligned}$$

The system combines these probabilities across publications and accumulates a composite distribution over the filtered population:

$$X_i = \sum\nolimits_{filter} Prob(seeing\ i)$$

where the summation Σ_{filter} may be weighted to a parent population if this is required.

Step 3.2 The system also estimates the actual (headcount) readerships as:

$$R_m = \sum\nolimits_{filter} \delta_m$$

and the cross-readerships as:

$$R_{m_1 m_2} = \sum\nolimits_{filter} \delta_{m1} \delta_{m2}$$

and uses these to estimate the mean, M, and variance, V,

of the true OTS distribution for the schedule, viz.,

$$M = R/P,$$

where R = total impacts $= \Sigma_m n_m R_m$, and

P = size of the filtered population;

$$V = S/P - M^2$$

where

$$S = R\,(2\textstyle\sum_m n_m - 1) - \sum_{m1 \neq m2} n_{m1} n_{m2} (R_{m1} + R_{m2} - R_{m1m2}) - \sum_m R_m n_m (n_m - 1)\,(1 + \gamma_m (P - R_m)/P)$$

γ_m = the casualness of publication m over the filtered population.

Again, the summations Σ_{filter} may be weighted to a parent population if this is required.

<u>Step 4</u> The system then modifies the composite distribution formed in step 3.1 so that its mean and variance match those of the true OTS distribution estimated in step 3.2, thus generating the final composite OTS distribution for each schedule. In this step, the reach (number of persons seeing at least one insertion) and average frequency (average impacts amongst those reached) can also be determined.

There are many ways of performing this step, and the system can utilize either of two options:

Alternative 4.1 Determining and then using multiplicative adjustment factors f_i (i = 0,..., N) which maximise the function:

$$\sum_{r=0}^{N} X_r \log(f_r X_r)$$

subject to the constraints:

$$\sum_{r=0}^{N} f_r X_r = P$$

$$\sum_{r=0}^{N} r.f_r X_r = R$$

$$\sum_{r=0}^{N} r^2.f_r X_r = S.$$

The solution to this problem is of the form:

$$f_r = \frac{1}{1 + \mu(r-R/P) + \nu(r^2-S/P)}$$

where μ and ν are (Lagrange multiplier) constants.

Alternative 4.2 Determining and then using absolute adjustments D_i ($i = 0,\ldots, N$) which minimise the weighted sum of squares

$$\sum_{r=0}^{N} w_r D_r^2$$

subject to the constraints:

$$\sum_{r=0}^{N} D_r = 0$$

$$\sum_{r=0}^{N} r.D_r = Error\ in\ R$$
$$= \Delta R$$

$$\sum_{r=0}^{N} r^2.D_r = Error\ in\ S$$
$$= \Delta S$$

The solution to this problem is given explicitly as:

$$D_r = \frac{1}{w_r}(1\ r\ r^2)\begin{pmatrix} \sum w_i^{-1} & \sum i\ w_i^{-1} & \sum i^2 w_i^{-1} \\ \sum i\ w_i^{-1} & \sum i^2 w_i^{-1} & \sum i^3 w_i^{-1} \\ \sum i^2 w_i^{-1} & \sum i^3 w_i^{-1} & \sum i^4 w_i^{-1} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \Delta R \\ \Delta S \end{pmatrix}$$

when $w_0 > 0$ and, as:

$$D_r = \frac{1}{w_r}(r\ r^2)\begin{pmatrix} \sum i^2 w_i^{-1} & \sum i^3 w_i^{-1} \\ \sum i^3 w_i^{-1} & \sum i^4 w_i^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \Delta R \\ \Delta S \end{pmatrix} \quad for\ r > 0$$

$$with\ D_0 = -\sum_{r=1}^{N} D_r$$

when $w_0 = 0$.

The formulae set out in steps 1-4 above cover the essential principles underlying the system. Slight variations may be introduced for average issue readerships of daily newspapers, for time dependent cross-readerships and for alternative definitions of regularity.

The weights w_i ($i = 0,..., N$) are weights used by the system in guiding the adjustment process. Where w_i is large the adjustment D_i to the corresponding X_i will be relatively small and vice versa. These weights are not to be confused with the respondent weights previously described.

N is the maximum number of insertions which it is possible for any respondent to see. This may be less than the total number of insertions in the schedule. A schedule may contain two or more media which are not available to the same respondents, for instance because they do not overlap geographically, so that no respondent can be exposed to all of them.

The first method of adjustment (4.1) involves smaller percentage changes in the distribution, but it does not guarantee additivity of the distribution over subsets of the population.

For a given set of weights $\underline{w}$, where N is constant for the subgroups and for the total group, the second method (4.2) does guarantee additivity. The second method also has a further advantage. Because it is possible to determine the errors in R and S, and hence the D_i, from summary statistics, this method can be used to estimate reach, and hence average frequency, by taking the complement of the OTS zero frequency (X_0) without having to determine the full OTS distribution. In cases where there are many insertions, this can save considerable computer time. For the same reason, this method can also be used to generate partial distributions (X_i) containing any number of terms $n \leq N$. For a given set of weights $\underline{w}$, the

answers are always consistent.

The disadvantage of the second method is that it can produce negative estimates. However, these are rare, and when they do occur, they are very small. Such estimates are analogous to the situation of negative counts that can be produced in some automatic counting machines through the subtraction of background noise.

The system almost completely overcomes the problem of declining reach and is not order dependent. It can model the actual shape of the OTS distribution. Its estimates of reach and frequency and its OTS distribution are always compatible with the headcount readership estimates. It does not dilute cross-media readerships.

It can also handle non-overlapping readerships. Indeed regions can be specifically excluded from consideration of coverage by selected media by having the appropriate respondent loyalties set to zero.

Clearly, however, the system is dependent upon the accuracy of the estimation/measurement of the respondent loyalties contained in the file (SURVEY).PRB, although not to the same extent as personal probability methods. For consistency and for best operation of the system, not only should:

$$\sum_{pop} p_m \approx R_m$$

but also the estimate of casualness determined by aggregating the individual beta distributions, viz.,

$$\gamma_m = G_m \sum_{pop} p_m (1-p_m) / [\bar{p}(1-\bar{p})]$$

should reconcile with the overall loyalty, L, determined

by averaging the loyalties of the actual readers, *i.e.*:

$$\gamma_m = \frac{1-L}{1-R_m/P}$$

$$\text{where} \quad L = \sum_{pop} p_m \delta_m / R_m.$$

It would be within the capacity of the skilled addressee to write computer code to achieve the methodology of the steps given above when guided by the flow diagram of Figure 1. In the alternative, the ASTEROID code may be obtained from the applicant company.

The steps of Figure 1 are well annotated and will not be discussed in detail. Common elements can be identified in the steps 1-4 above. Even so, the following comments will be made with reference to Figure 1.

Data are stored as binary (bit) strings or, in the case of the 'loyalty data' as a character string, with each bit/character representing one respondent. Because the sample size is usually large and there are memory constraints, the sample is usually broken up into 'segments'. Each segment is processed completely, and the results are cumulated before moving to the next segment.

Data extracted from the survey database in respect of one possible response (*e.g.* 'read the newspaper') are thus in the form of a bit vector representing the segment concerned, with 'on' bits indicating, in this instance, respondents claiming to have read the newspaper and 'off' bits those claiming not to have read it. In some cases bit vectors are combined to produce real number vectors (*e.g.* a person claiming to have read three issues of a daily paper out of a possible five issues would be given a value of 0.6). At completion of the calculations, the

results are tabulated and output to a device such as a printer, hard disk file or video screen. For the schedule and filters described earlier, the composite OTS distribution, reach and average frequency are as follows. The abbreviation, 'Cume.' represents a cumulative representation of the composite OTS distributions ('Dist.').

Schedule File: SCH01
Filter: Gentlemen
Population (000's): 551.2

Publications, m	Schedules (Insertions, n)			
NAT-GEOGRAPH	3	4	2	4
NEW-IDEA	2	3	2	2
ADV-M-F	1	2	2	0

Reach (000's):	307.3	349.1	345.5	105.2
(%):	55.8%	63.3%	62.7%	19.1%
Ave. frequency:	1.54	2.36	2.04	2.30
Impacts (000's):	472.7	824.4	703.3	242.1

O.T.S.	Dist.	Cume.	Dist.	Cume.	Dist.	Cume.	Dist.	Cume.
0	44.2%		36.7%		37.3%		80.9%	
1	39.4%	55.8%	15.5%	63.3%	16.7%	62.7%	6.4%	19.1%
2	7.2%	16.3%	31.5%	47.8%	34.7%	45.9%	6.4%	12.7%
3	5.6%	9.1%	5.7%	16.3%	5.1%	11.2%	1.9%	6.3%
4	2.7%	3.5%	4.2%	10.7%	5.0%	6.2%	3.5%	4.4%
5	0.6%	0.8%	3.4%	6.5%	0.7%	1.1%	0.5%	0.9%
6	0.2%	0.2%	2.1%	3.1%	0.4%	0.4%	0.4%	0.4%
7			0.5%	1.0%				
8			0.3%	0.5%				
9			0.2%	0.2%				

Schedule File: SCH01

Filter: Ladies

Population (000's): 570.8

Publications, m	Schedules Insertions, n)			
NAT-GEOGRAPH	3	4	2	4
NEW-IDEA	2	3	2	2
ADV-M-F	1	2	2	0

Reach (000's):	387.0	424.5	410.5	236.9
(%):	67.8%	74.4%	71.9%	41.5%
Ave. frequency	1.82	2.76	2.28	1.97
Impacts (000's):	702.5	1171.1	937.2	467.8

O.T.S.	Dist.	Cume.	Dist.	Cume.	Dist.	Cume.	Dist.	Cume
0	32.2%		25.6%		28.1%		58.5%	
1	35.1%	67.8%	15.3%	74.4%	16.5%	71.9%	14.6%	41.5%
2	17.6%	32.7%	26.7%	59.1%	34.6%	55.4%	20.4%	26.9%
3	10.5%	15.0%	14.0%	32.4%	8.6%	20.7%	2.1%	6.5%
4	2.5%	4.6%	6.6%	18.4%	9.3%	12.1%	2.6%	4.4%
5	1.2%	2.1%	7.3%	11.8%	1.5%	2.8%	0.9%	1.8%
6	0.9%	0.9%	1.9%	4.5%	1.3%	1.3%	0.9%	0.9%
7			1.0%	2.5%				
8			0.7%	1.5%				
9			0.8%	0.8%				

Schedule File: SCH01
Filter: All Adults
Population (000's): 1122.0

Publications, m	Schedules (Insertions, n)			
NAT-GEOGRAPH	3	4	2	4
NEW-IDEA	2	3	2	2
ADV-M-F	1	2	2	0

Reach (000's):	694.3	773.6	756.0	342.1
(%):	61.9%	69.0%	67.4%	30.5%
Ave. frequency:	1.69	2.58	2.17	2.08
Impacts (000's)	1175.2	1995.5	1640.5	709.9

O.T.S.	Dist.	Cume.	Dist.	Cume.	Dist.	Cume.	Dist.	Cume.
0	38.1%		31.0%		32.6%		69.5%	
1	37.3%	61.9%	15.4%	69.0%	16.7%	67.4%	10.6%	30.5%
2	12.5%	24.6%	29.1%	53.6%	34.7%	50.7%	13.5%	19.9%
3	8.1%	12.1%	9.9%	24.5%	6.9%	16.1%	2.0%	6.4%
4	2.6%	4.0%	5.4%	14.6%	7.2%	9.2%	3.0%	4.4%
5	0.9%	1.5%	5.4%	9.2%	1.1%	2.0%	0.7%	1.4%
6	0.6%	0.6%	2.0%	3.8%	0.9%	0.9%	0.7%	0.7%
7			0.8%	1.8%				
8			0.5%	1.0%				
9			0.5%	0.5%				

From the foregoing tables, users can determine which schedule best meets their objectives, and an assessment of the worth of various advertising strategies can be made.

The invention described above may be embodied in other specific forms from those discussed. The embodiments described are to be considered as illustrative and not limiting on the scope of the invention.

CLAIMS:

1. In an information processing system including a processor means, a memory means coupled to the processor means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means, an estimation method comprising the steps of:

(A) selecting, by the input means, at least one schedule of media vehicles (m) and, for each selected schedule, specifying a number of insertions (n_m) in each media vehicle;

(B) defining, by the input means, at least one filter specifying a set of the survey respondents for which each schedule is to be evaluated;

(C) applying, by the processor means, the filter to the database;

(D) for each media vehicle, calculating, by the processor means, an array of the probabilities of being exposed to i out of n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle;

(E) combining, by the processor means, the data relating to the usage of each selected media vehicle with said array of probabilities to give a probability of being exposed to i insertions in each selected media vehicle,

combining the probabilities of being exposed to i insertions across the media vehicles and accumulating over all respondents passing the filter to yield a composite distribution (X_i);

(F) by the processor means, within the set of survey respondents defined by the filter:

(i) summing media vehicle usage to give total estimated usage for each selected media vehicle, and summing the media vehicle usage of each pair of media vehicles (m_1, m_2) to give total estimated cross-usage of each pair of media vehicles,

(ii) estimating the mean of the true OTS (opportunities-to-see) distribution for each selected schedule from said total estimated usage for each selected media vehicle,

(iii) estimating the variance of the true OTS distribution for each schedule from the said mean, the said total estimated cross-usage and the said estimates of regularity of exposure for each media vehicle;

(G) operating, by the processor means, on each said composite distribution (X_i) to modify it so that it matches with the said mean and the said variance of the true OTS distribution to form a final composite OTS distribution; and

(H) outputting, by the output means, the final composite OTS distribution for each selected schedule.

2. In an information processing system including a processor means, a memory means coupled to the processor

means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means, an estimation method comprising the steps of:

(A) inputting, by the input means, at least one schedule of media vehicles (m) and, for each schedule, specifying a number of insertions (n_m) in each media vehicle;

(B) generating, by the processor means, a composite OTS distribution (X_i);

(C) operating, by the processor means, on each said composite OTS distribution (X_i) to modify it so that it matches with the mean and the variance of the true OTS distribution as estimated from media vehicle usage and regularity of exposure to form a final composite OTS distribution; and

(D) outputting, by the output means, the final composite OTS distributions.

3. In an information processing system including a processor means, a memory means coupled to the processor means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means, an estimation method comprising the steps of:

(A) inputting, by the input means, at least one schedule of media vehicles (m) and, for each schedule, specifying a number of insertions (n_m) in each media vehicle;

(B) for each media vehicle calculating, by the processor means, a distribution array of the probabilities of being exposed to i out of n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle;

(C) utilising these distribution arrays in the generation, by the processor means, of a composite OTS (opportunities-to-see) distribution for each schedule; and

(D) outputting, by the output means, the composite OTS distributions.

4. In an information processing system including a processor means, a memory means coupled to the processor means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means, an estimation method comprising the steps of:

(A) selecting, by the input means, at least one schedule of media vehicles (m) and, for each selected schedule, specifying a number of insertions (n_m) in each media vehicle;

(B) defining, by the input means, at least one filter specifying a set of the survey respondents for

which each schedule is to be evaluated;

(C) applying, by the processor means, the filter to the database;

(D) for each media vehicle, calculating, by the processor means, a partial array of the probabilities of being exposed to some numbers i of the possible n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle;

(E) combining, by the processor means, the data relating to the usage of each selected media vehicle with said partial array of probabilities to give a probability of being exposed to i insertions in each selected media vehicle, combining the probabilities of being exposed to i insertions across the media vehicles and accumulating over all respondents passing the filter to yield a partial composite distribution (X_i);

(F) by the processor means, within the set of survey respondents defined by the filter:

(i) summing media vehicle usage to give total estimated usage for each selected media vehicle, and summing the media vehicle usage of each pair of media vehicles (m_1, m_2) to give total estimated cross-usage of each pair of media vehicles,

(ii) estimating the mean of the true OTS (opportunities-to-see) distribution for each selected schedule from said total estimated usage for each selected media vehicle,

(iii) estimating the variance of the true OTS distribution for each schedule from the said mean, the said total estimated cross-usage and the said estimates of regularity of exposure for each media vehicle;

(G) operating, by the processor means, on each said partial composite distribution (X_i) to modify it so that it matches with the said mean and the said variance of the true OTS distribution to form a final partial composite OTS distribution; and

(H) outputting, by the output means, the final partial composite OTS distribution for each selected schedule.

5. In an information processing system including a processor means, a memory means coupled to the processor means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means, an estimation method comprising the steps of:

(A) selecting, by the input means, at least one schedule of media vehicles (m) and, for each selected schedule, specifying a number of insertions (n_m) in each media vehicle;

(B) defining, by the input means, at least one filter specifying a set of the survey respondents for which each schedule is to be evaluated;

(C) applying, by the processor means, the filter to

the database;

(D) for each media vehicle, calculating, by the processor means, a partial array of the probabilities of being exposed to some numbers i of the possible n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle;

(E) combining, by the processor means, the data relating to the usage of each selected media vehicle with said partial array of probabilities to give a probability of being exposed to zero insertions in each selected media vehicle, combining the probabilities of being exposed to zero insertions across the media vehicles and accumulating over all respondents passing the filter to yield a composite distribution frequency (X_0) of zero exposures;

(F) by the processor means, within the set of survey respondents defined by the filter:

(i) summing media vehicle usage to give total estimated usage for each selected media vehicle, and summing the media vehicle usage of each pair of media vehicles (m_1, m_2) to give total estimated cross-usage of each pair of media vehicles,

(ii) estimating the mean of the true OTS (opportunities-to-see) distribution for each selected schedule from said total estimated usage for each selected media vehicle,

(iii) estimating the variance of the true OTS distribution for each schedule from the said mean,

the said total estimated cross-usage and the said estimates of regularity of exposure for each media vehicle;

(G) operating, by the processor means, on each said composite distribution frequency (X_0) of zero exposures to modify it so that it matches with the said mean and the said variance of the true OTS distribution to yield a final composite OTS distribution frequency of zero exposures and therefore the reach; and

(H) outputting, by the output means, the reach for each selected schedule.

6. The method of any one of claims 1-3, comprising the further step of calculating, by the processor means, the reach for each selected schedule by forming the sum of the final composite OTS distribution frequencies for the cases of the number of insertions being one and more.

7. The method of claim 6, comprising the further step of calculating, by the processor means, the average frequency for each selected schedule determined as the total impacts divided by the reach.

8. The method of any one of claims 1-4, comprising the further step of calculating, by the processor means, the reach for each selected schedule by taking the complement of the final composite OTS distribution frequency for the case of the number of insertions being zero.

9. The method of claim 8, comprising the further step of calculating, by the processor means, the average frequency for each selected schedule determined as the total impacts divided by the reach.

10. The method of either one of claims 1 or 2, wherein the step (G) comprises use of multiplicative adjustment factors (f_0, $f_1, \ldots,$ f_N) which maximise a summation function involving said composite distribution (X_0, $X_1, \ldots,$ X_N).

11. The method of claim 10, wherein the summation function is:

$$\sum_{r=0}^{N} X_r \log(f_r X_r).$$

12. The method of any one of claims 1,2,4 or 5, wherein the step (G) comprises use of absolute adjustments, (D_0, $D_1, \ldots,$ D_N) which minimise a weighted sum of squares,

$$\sum_{i=0}^{N} w_i D_i^2.$$

13. The method of any one of claims 1-5, comprising the further step of:

(I) printing, by the output means, the final output composite OTS distribution for each schedule.

14. The method of any one of claims 1-5, comprising the further steps before step (A) of conducting a survey to obtain the said database of survey responses, and storing the database in the said memory means.

15. The method of claim 14, comprising the further step of:

(I) printing, by the output means, the final output composite OTS distribution for each schedule.

16. The method of any one of claims 1-5, comprising the further steps of:

(AA) determining, by the processor means, whether any adjustment is to be made to allow for

(a) media vehicles not included in the survey throughout the whole survey period;
(b) sample subsets defined wholly or partly in terms of time periods;
(c) sample subsets defined in terms of information not collected throughout the whole survey period;
(d) any sample subset used at any stage in the calculation which is formed as the intersection of any two or more sample subsets of the types (a)-(c);

(BB) calculating, by the processor means, the adjustment factors for

(a) the total estimated usage of each media vehicle;
(b) the total estimated cross-usage of each pair of media vehicles;
(c) the frequencies (X_0, X_1,..., X_N) of the composite frequency distribution; and

(CC) applying, by the processor means, these adjustment factors to produce revised estimates of the composite distribution frequencies, total impacts and the mean and variance of the true OTS distribution.

17. The method of claim 16, wherein the said adjustment factors are calculated by dividing a weighted number of respondents in the survey by a weighted number of respondents providing information within the identified

time period(s) or intersections of time periods.

18. The method of claim 15, comprising the further steps of:

(AA) determining, by the processor means, whether any adjustment is to be made to allow for

(a) media vehicles not included in the survey throughout the whole survey period;
(b) sample subsets defined wholly or partly in terms of time periods;
(c) sample subsets defined in terms of information not collected throughout the whole survey period;
(d) any sample subset used at any stage in the calculation which is formed as the intersection of any two or more sample subsets of the types (a)-(c);

(BB) calculating, by the processor means, the adjustment factors for

(a) the total estimated usage of each media vehicle;
(b) the total estimated cross-usage of each pair of media vehicles;
(c) the frequencies (X_0, X_1,..., X_N) of the composite frequency distribution; and

(CC) applying, by the processor means, these adjustment factors to produce revised estimates of the composite distribution frequencies, total impacts and the mean and variance of the true OTS distribution.

19. In an information processing system including a processor means, a memory means coupled to the processor

means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means, an estimation method for determining a composite OTS (opportunities-to-see) distribution for at least one schedule of media vehicles, with a specified number of insertions in each media vehicle and at least one filter specifying a set of survey respondents for which each schedule is to be evaluated, the method comprising the steps of:

(AA) determining, by the processor means, whether any adjustment is to be made to allow for

(a) media vehicles not included in the survey throughout the whole survey period;
(b) sample subsets defined wholly or partly in terms of time periods;
(c) sample subsets defined in terms of information not collected throughout the whole survey period;
(d) any sample subset used at any stage in the calculation which is formed as the intersection of any two or more sample subsets of the types (a)-(c);

(BB) calculating, by the processor means, the adjustment factors for

(a) the total estimated usage of each media vehicle;
(b) the total estimated cross-usage of each pair of media vehicles;

(c) the frequencies (X_0, X_1,..., X_N) of the composite frequency distribution; and

(CC) applying, by the processor means, these adjustment factors to produce revised estimates of the composite distribution frequencies, total impacts and the mean and variance of the true OTS distribution.

20. The method of claim 19, wherein the said adjustment factors are calculated by dividing a weighted number of respondents in the survey by a weighted number of respondents providing information within the identified time period(s) or intersections of time periods.

21. In an information processing system including a processor means, a memory means coupled to the processor means and operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles, input means coupled to the processor means and operable for obtaining user input, and output means coupled to the processor means, an estimation method comprising the steps of:

(A) inputting, by the input means, at least one schedule of media vehicles (m) and, for each schedule, specifying a number of insertions (n_m) in each media vehicle;

(B) generating, by the processor means, a distribution array of the probabilities of being exposed to i out of n_m insertions using non-binomial distributions for each respondent;

(C) utilising the distribution array in the generation, by the processor means, of a composite OTS

(opportunities-to-see) distribution for each schedule; and

(D) outputting, by the output means, the composite OTS distributions.

22. An information processing system for use in media research, the system comprising:

(A) a memory means operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles;

(B) input means operable for obtaining user input and selection of at least one schedule of media vehicles (m) and, for each selected schedule, specification of a number of insertions (n_m) in each media vehicle and at least one filter specifying a set of the survey respondents for which each schedule is to be evaluated;

(C) a processor means coupled to the memory means and the input means for applying the filter to the database, and for each media vehicle, calculating an array of the probabilities of being exposed to i out of n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle, for combining the data relating to the usage of each selected media vehicle with said array of probabilities to give a probability of being exposed to i insertions in each selected media vehicle, combining the probabilities of being exposed to i insertions across the media vehicles and accumulating over all respondents passing the filter to yield a composite distribution (X_i), and within the set of survey respondents defined by the filter:

(i) summing media vehicle usage to give total estimated usage for each selected media vehicle, and summing the media vehicle usage of each pair of media vehicles (m_1, m_2) to give total estimated cross-usage of each pair of media vehicles,

(ii) estimating the mean of the true OTS (opportunities-to-see) distribution for each selected schedule from said total estimated usage for each selected media vehicle,

(iii) estimating the variance of the true OTS distribution for each schedule from the said mean, the said total estimated cross-usage and the said estimates of regularity of exposure for each media vehicle,

operating on each said composite distribution (X_i) to modify it so that it matches with the said mean and the said variance of the true OTS distribution to form a final composite OTS distribution; and

(D) output means coupled to the processor means for outputting the final composite OTS distribution for each selected schedule.

23. An information processing system for use in media research, the system comprising:

(A) a memory means operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles;

(B) input means operable for obtaining user input and selection of at least one schedule of media vehicles

(m) and, for each selected schedule, specification of a number of insertions (n_m) in each media vehicle and at least one filter specifying a set of the survey respondents for which each schedule is to be evaluated;

(C) a processor means coupled to the memory means and the input means for applying the filter to the database, and for each media vehicle, calculating a partial array of the probabilities of being exposed to some numbers i of the possible n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle, for combining the data relating to the usage of each selected media vehicle with said partial array of probabilities to give a probability of being exposed to i insertions in each selected media vehicle, combining the probabilities of being exposed to i insertions across the media vehicles and accumulating over all respondents passing the filter to yield a partial composite distribution (X_i), and within the set of survey respondents defined by the filter:

(i) summing media vehicle usage to give total estimated usage for each selected media vehicle, and summing the media vehicle usage of each pair of media vehicles (m_1, m_2) to give total estimated cross-usage of each pair of media vehicles,

(ii) estimating the mean of the true OTS (opportunities-to-see) distribution for each selected schedule from said total estimated usage for each selected media vehicle,

(iii) estimating the variance of the true OTS distribution for each schedule from the said mean, the said total estimated cross-usage and the said

estimates of regularity of exposure for each media vehicle,

operating on each said partial composite distribution (X_i) to modify it so that it matches with the said mean and the said variance of the true OTS distribution to form a final partial composite OTS distribution; and

(D) output means coupled to the processor means for outputting the final partial composite OTS distribution for each selected schedule.

24. An information processing system for use in media research, the system comprising:

(A) a memory means operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles;

(B) input means operable for obtaining user input and selection of at least one schedule of media vehicles (m) and, for each selected schedule, specification of a number of insertions (n_m) in each media vehicle and at least one filter specifying a set of the survey respondents for which each schedule is to be evaluated;

(C) a processor means coupled to the memory means and the input means for applying the filter to the database, and for each media vehicle, calculating a partial array of the probabilities of being exposed to some numbers i of the possible n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicle, for combining the data relating to the usage of each selected media vehicle with

said partial array of probabilities to give a probability of being exposed to zero insertions in each selected media vehicle, combining the probabilities of being exposed to zero insertions across the media vehicles and accumulating over all respondents passing the filter to yield a composite distribution frequency of zero exposures (X_0), and within the set of survey respondents defined by the filter:

(i) summing media vehicle usage to give total estimated usage for each selected media vehicle, and summing the media vehicle usage of each pair of media vehicles (m_1, m_2) to give total estimated cross-usage of each pair of media vehicles,

(ii) estimating the mean of the true OTS (opportunities-to-see) distribution for each selected schedule from said total estimated usage for each selected media vehicle,

(iii) estimating the variance of the true OTS distribution for each schedule from the said mean, the said total estimated cross-usage and the said estimates of regularity of exposure for each media vehicle,

operating on each said composite distribution frequency of zero exposures (X_0) to modify it so that it matches with the said mean and the said variance of the true OTS distribution to form a final composite OTS distribution frequency of zero exposures and therefore, by subtraction, the reach; and

(D) output means coupled to the processor means for outputting the reach for each selected schedule.

25. An information processing system for media research, the system comprising:

(A) a memory means operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles;

(B) input means operable for obtaining user input and for selection of at least one schedule of media vehicles (m) and, for each selected schedule, specification of a number of insertions (n_m) in each media vehicle;

(C) a processor means coupled to the memory means and the input means for generating distribution arrays of the probabilities of being exposed to i out of n_m insertions using beta distributions for each respondent, said beta distributions having parameters that are a function of the regularity of exposure to the media vehicles, utilising the distribution arrays in the generation of a composite OTS (opportunities-to-see) distribution for each schedule; and

(D) output means coupled to the processor means for outputting the composite OTS distributions.

26. An information processing system for media research, the system comprising:

(A) a memory means operable for storing therein a database of survey responses including media vehicle usage and estimates of regularity of exposure of respondents to the media vehicles;

(B) input means operable for obtaining user input

and for selection of at least one schedule of media vehicles (m) and, for each selected schedule, specification of a number of insertions (n_m) in each media vehicle;

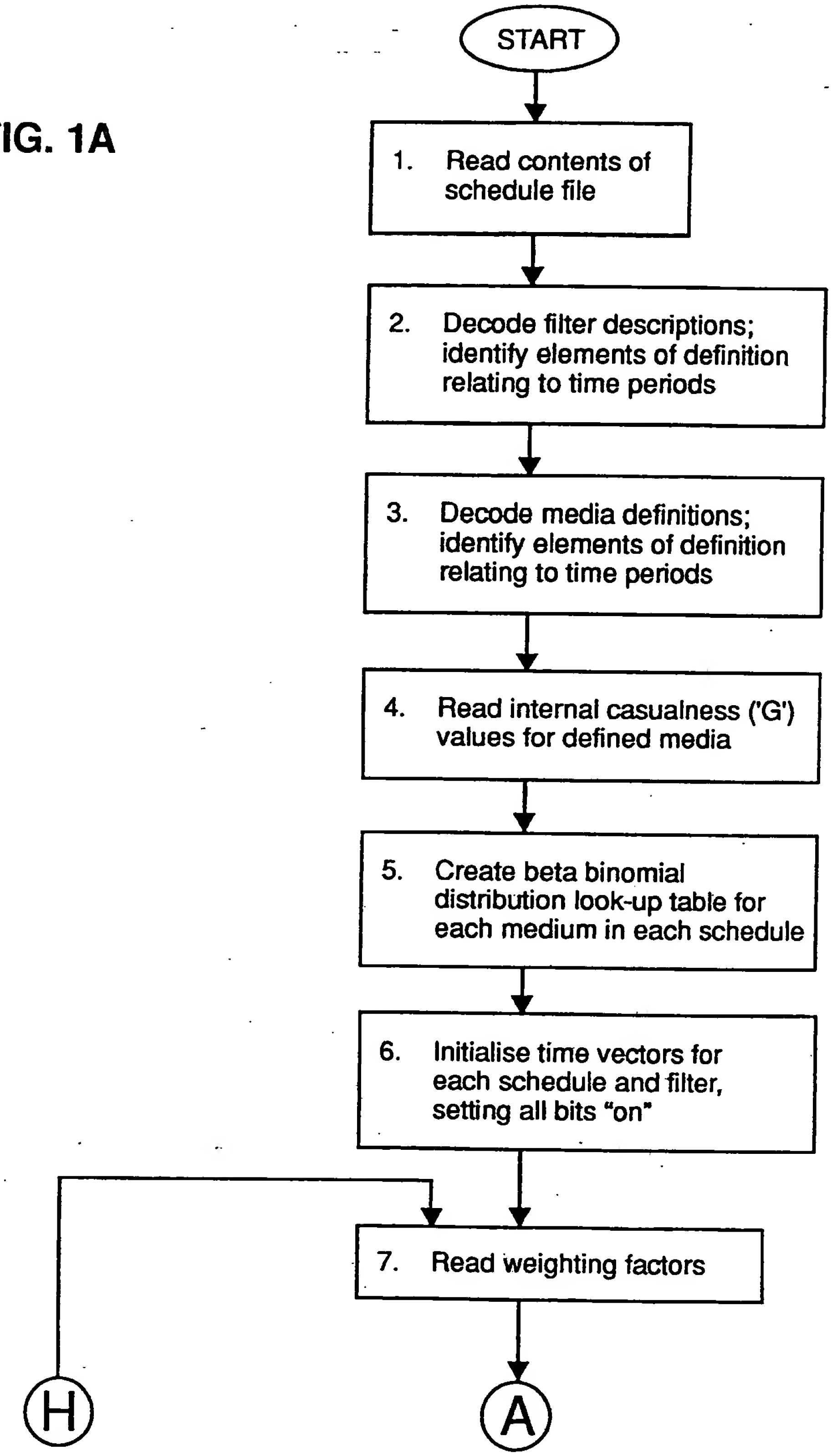
(C) a processor means coupled to the memory means and the input means for generating a composite OTS distribution (X_i), for operating on each said composite OTS distribution (X_i) to modify it so that it matches with the mean and the variance of the true OTS distribution as estimated from media vehicle usage and regularity of exposure to form a final composite OTS distribution for each schedule; and

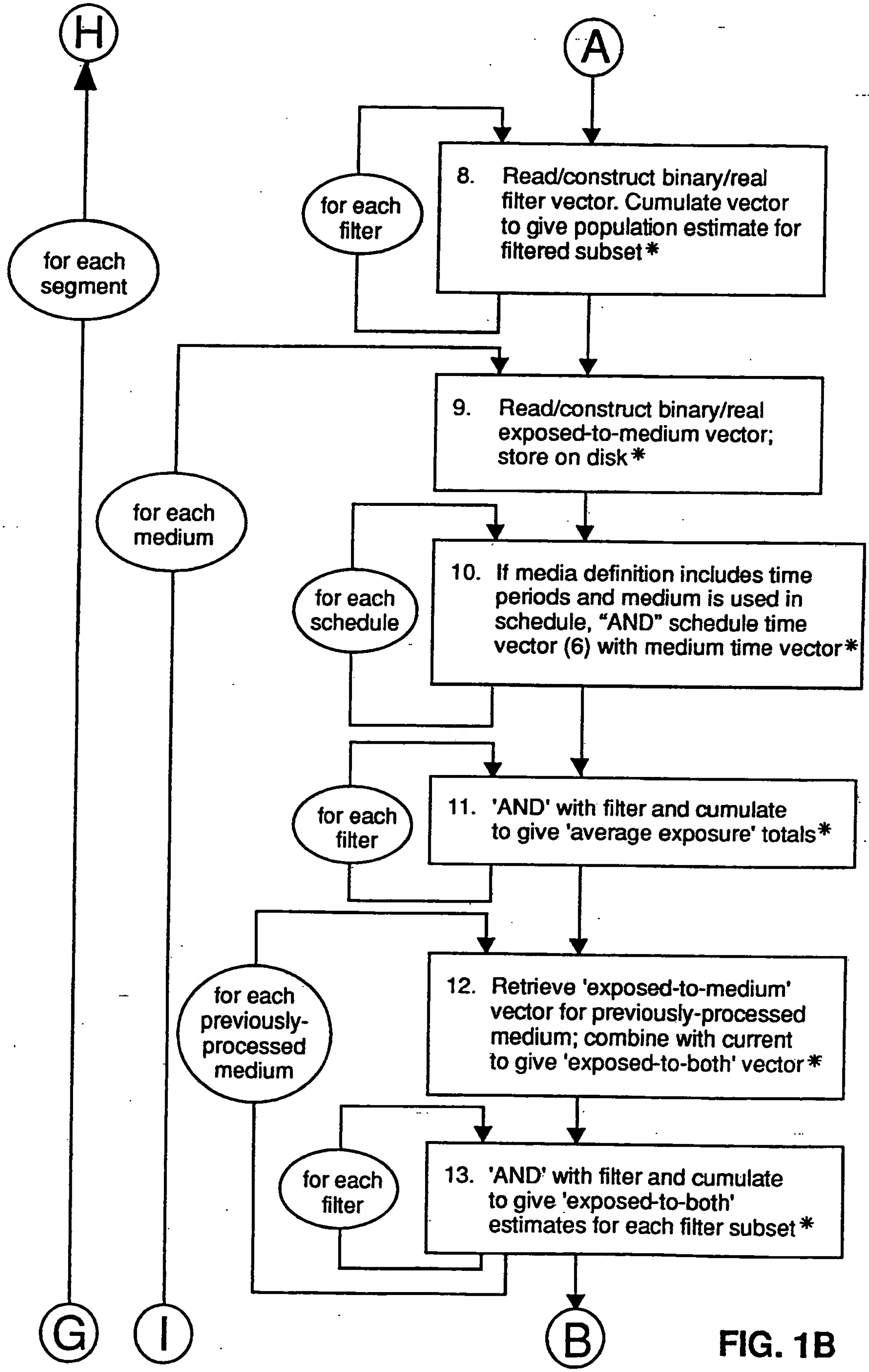
(D) output means coupled to the processor means for outputting the final composite OTS distributions.

27. The apparatus of any one of claims 22-26, further comprising:

(E) printer means for printing the final composite OTS distributions.

FIG. 1A


START
1. Read contents of schedule file
2. Decode filter descriptions; identify elements of definition relating to time periods
3. Decode media definitions; identify elements of definition relating to time periods
4. Read internal casualness ('G') values for defined media
5. Create beta binomial distribution look-up table for each medium in each schedule
6. Initialise time vectors for each schedule and filter, setting all bits "on"
7. Read weighting factors
H
A



H
A
for each segment
for each filter
8. Read/construct binary/real filter vector. Cumulate vector to give population estimate for filtered subset*
for each medium
9. Read/construct binary/real exposed-to-medium vector; store on disk*
for each schedule
10. If media definition includes time periods and medium is used in schedule, "AND" schedule time vector (6) with medium time vector*
for each filter
11. 'AND' with filter and cumulate to give 'average exposure' totals*
for each previously-processed medium
12. Retrieve 'exposed-to-medium' vector for previously-processed medium; combine with current to give 'exposed-to-both' vector*
for each filter
13. 'AND' with filter and cumulate to give 'exposed-to-both' estimates for each filter subset*
G
I
B

FIG. 1B

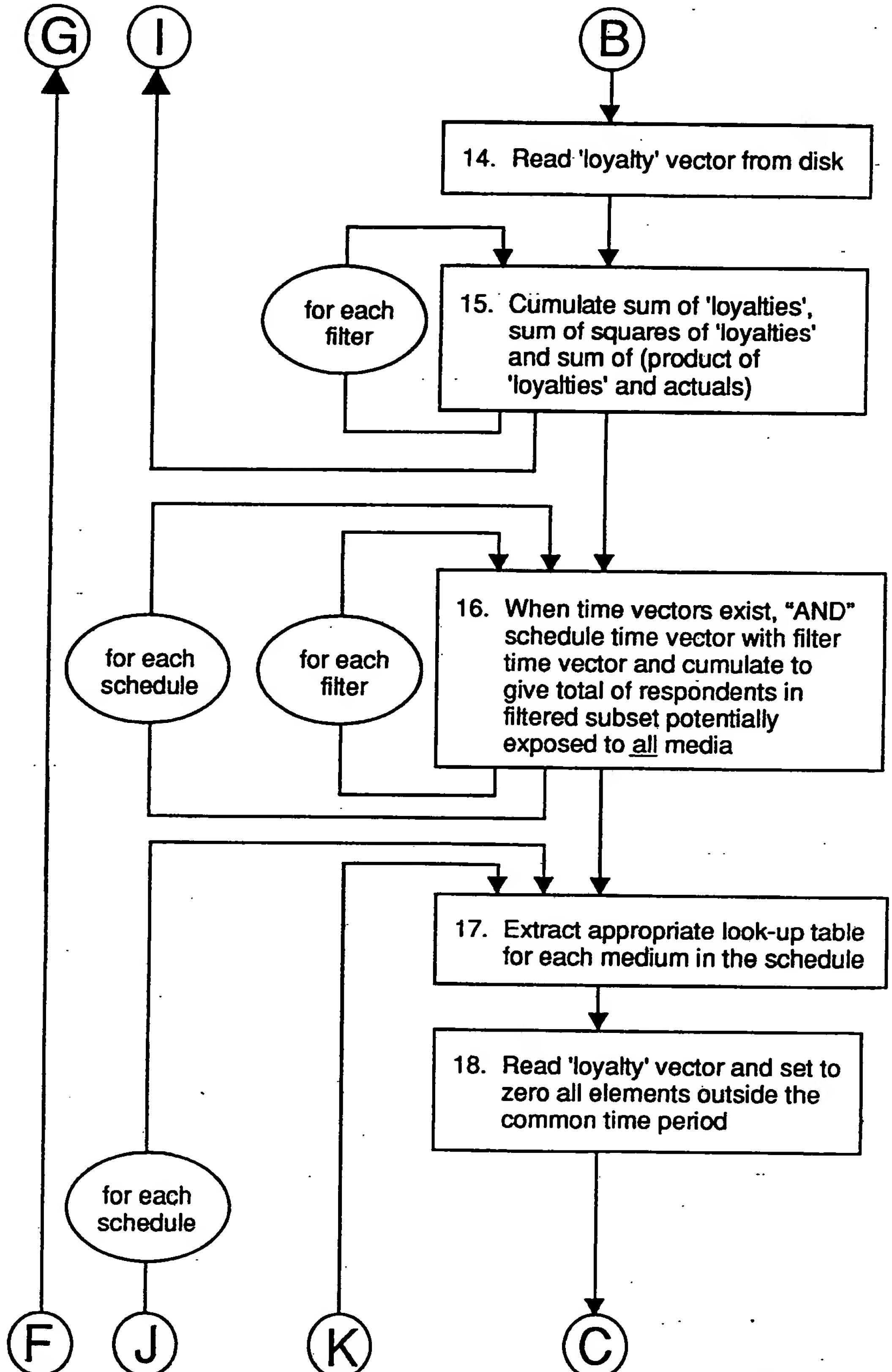

G
I
B
14. Read 'loyalty' vector from disk
for each filter
15. Cumulate sum of 'loyalties', sum of squares of 'loyalties' and sum of (product of 'loyalties' and actuals)
for each schedule
for each filter
16. When time vectors exist, "AND" schedule time vector with filter time vector and cumulate to give total of respondents in filtered subset potentially exposed to all media
17. Extract appropriate look-up table for each medium in the schedule
18. Read 'loyalty' vector and set to zero all elements outside the common time period
for each schedule
F
J
K
C

FIG. 1C

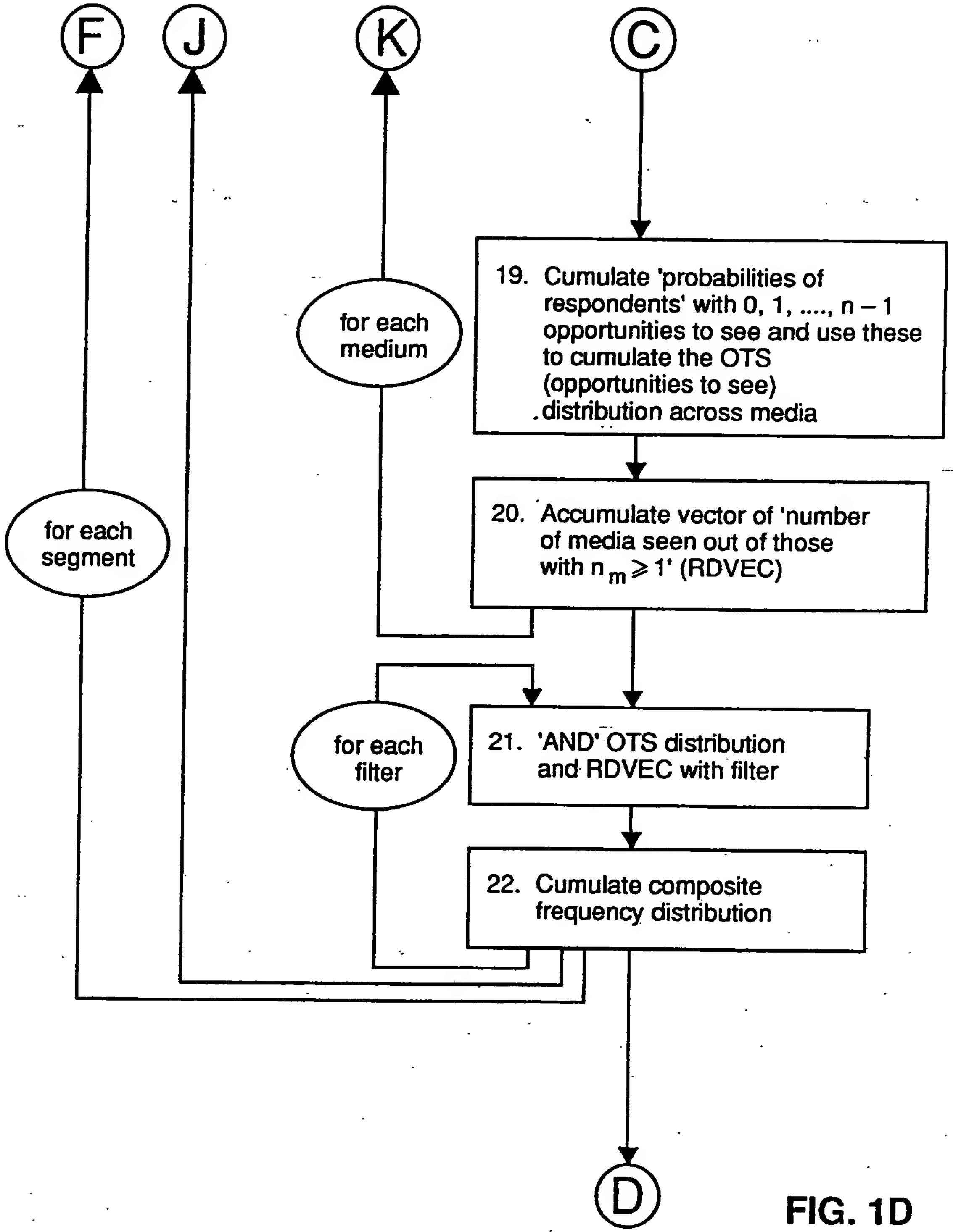

F
J
K
C
19. Cumulate 'probabilities of respondents' with 0, 1, ..., n – 1 opportunities to see and use these to cumulate the OTS (opportunities to see) distribution across media
for each medium
20. Accumulate vector of 'number of media seen out of those with $n_m \geqslant 1$' (RDVEC)
for each segment
for each filter
21. 'AND' OTS distribution and RDVEC with filter
22. Cumulate composite frequency distribution
D

FIG. 1D

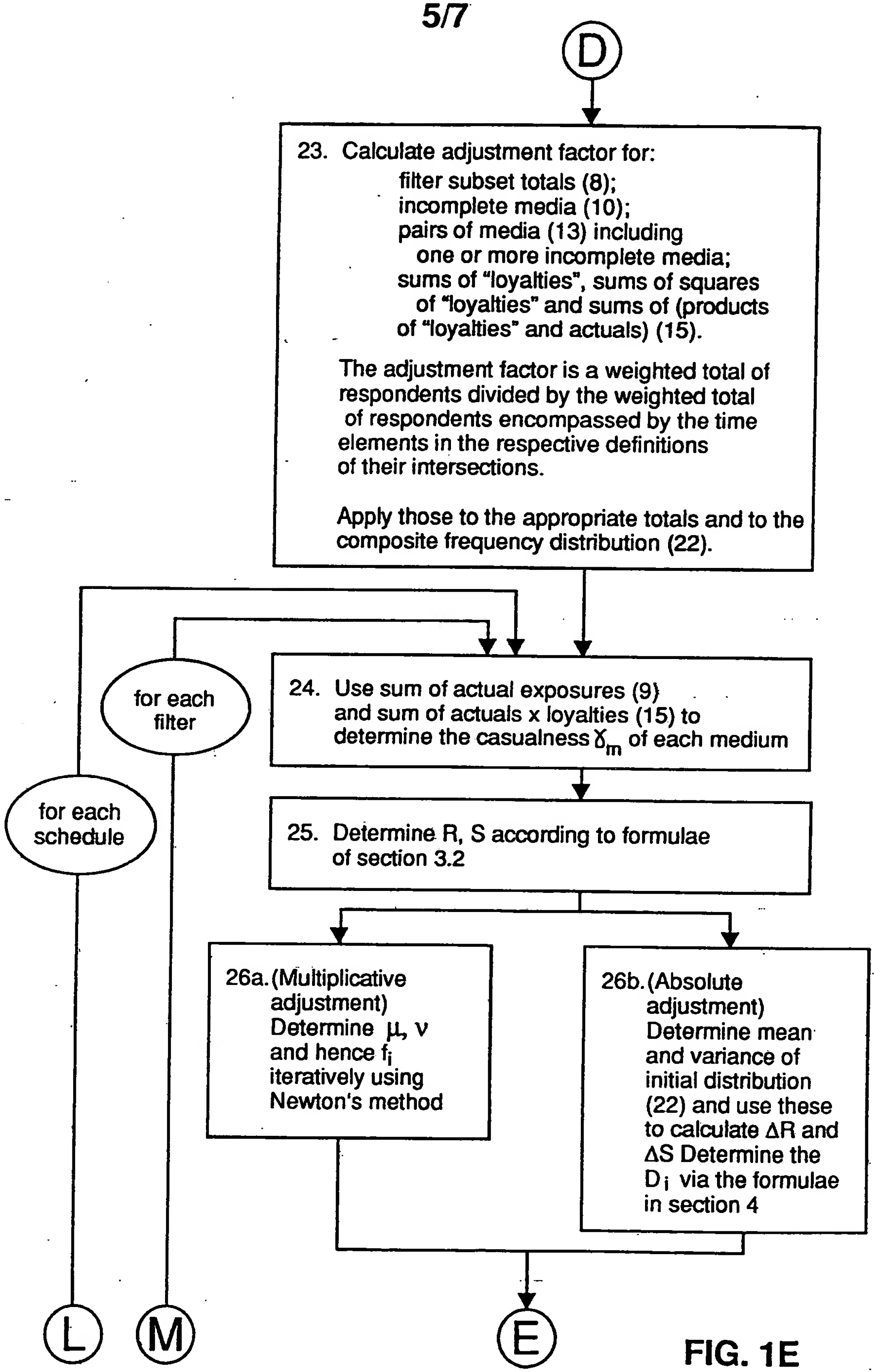
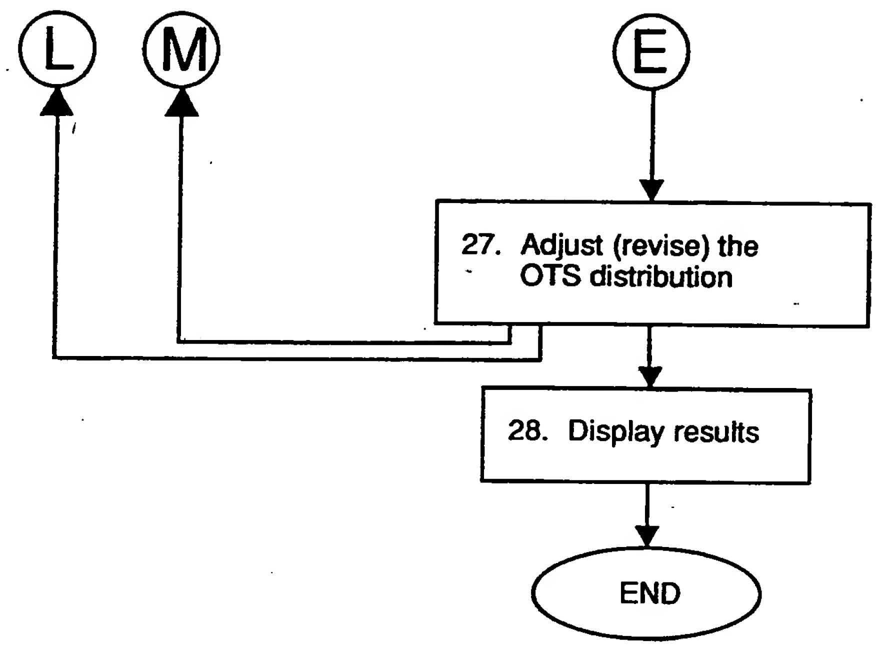

D
23. Calculate adjustment factor for:
filter subset totals (8);
incomplete media (10);
pairs of media (13) including one or more incomplete media;
sums of "loyalties", sums of squares of "loyalties" and sums of (products of "loyalties" and actuals) (15).
The adjustment factor is a weighted total of respondents divided by the weighted total of respondents encompassed by the time elements in the respective definitions of their intersections.
Apply those to the appropriate totals and to the composite frequency distribution (22).
for each filter
for each schedule
24. Use sum of actual exposures (9) and sum of actuals x loyalties (15) to determine the casualness γ_m of each medium
25. Determine R, S according to formulae of section 3.2
26a. (Multiplicative adjustment) Determine μ, ν and hence f_i iteratively using Newton's method
26b. (Absolute adjustment) Determine mean and variance of initial distribution (22) and use these to calculate ΔR and ΔS Determine the D_i via the formulae in section 4
L
M
E

FIG. 1E



* For each filter and medium comprising or including time elements in its definition a parallel "time" vector is constructed with "on" bits representing respondents interviewed in the time period(s) concerned. When these time vectors exist they are processed in the same way as the filter and medium vectors.

FIG. 1F

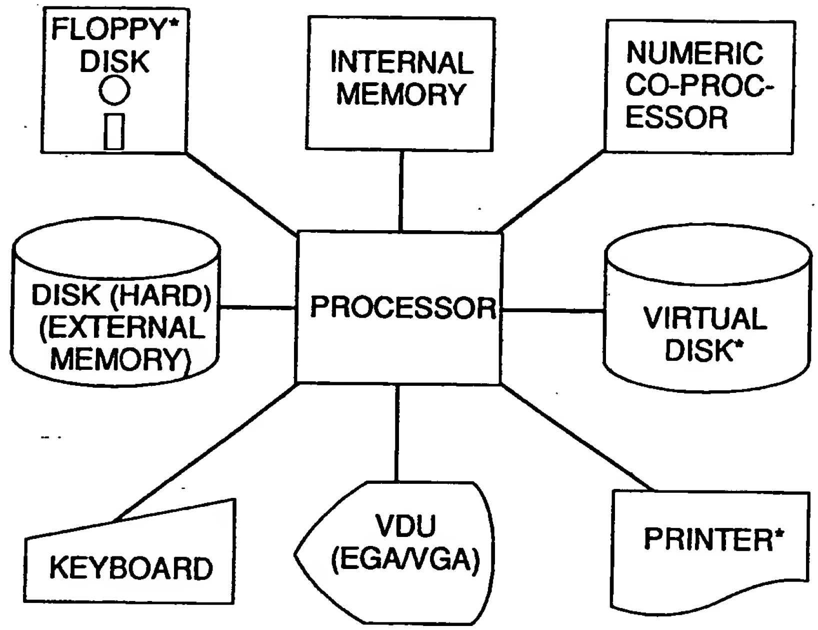

FLOPPY* DISK
INTERNAL MEMORY
NUMERIC CO-PROC-ESSOR
DISK (HARD) (EXTERNAL MEMORY)
PROCESSOR
VIRTUAL DISK*
KEYBOARD
VDU (EGA/VGA)
PRINTER*

FIG. 2

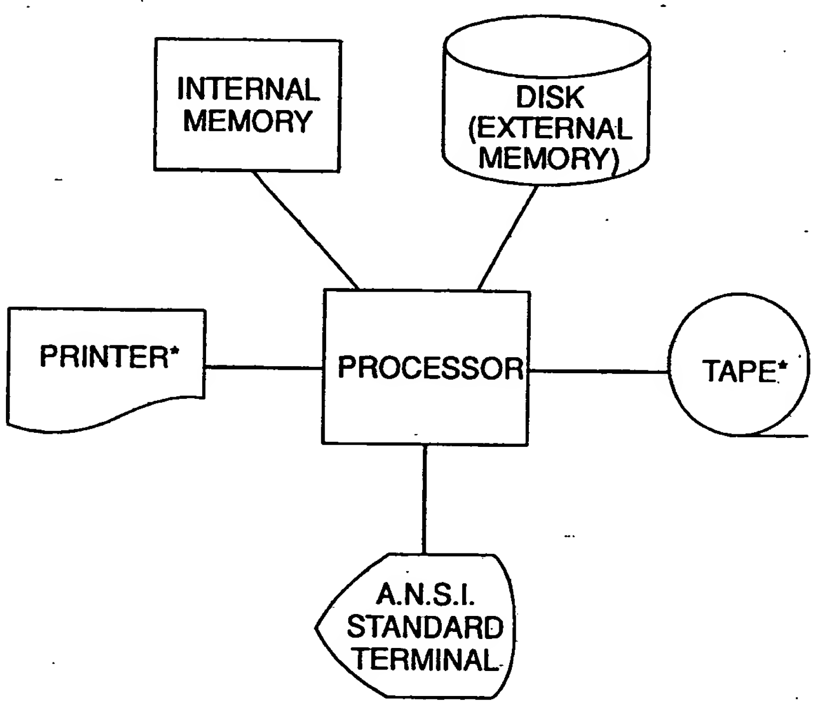

INTERNAL MEMORY
DISK (EXTERNAL MEMORY)
PRINTER*
PROCESSOR
TAPE*
A.N.S.I. STANDARD TERMINAL

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU92/00286

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.5 G06F 15/36

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC - G06F 15/36

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
AU : IPC as above

Electronic data base consulted during the international search (name of data base, and where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate of the relevant passages	Relevant to Claim No.
A	US,A, 4781 596 (Weinblatt) 1 November 1988 (1.11.88)	
A	US,A, 4659 314 (Weinblatt) 21 April 1987 (2	

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 August 1992 (31.08.92)	2 Oct 1992 (02.10.92)

Name and mailing address of the ISA/	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No. 06 2853929	Ed Knock Telephone No. (06) 2832206

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
US	4781596						
US	4659314	EP	231575	JP	62189824	US	4726771
		US	4939326	US	5019679		
							END OF ANNEX